Analyzing the Effect of Literal Death for Member Engagement in Online Health Community

Saumik Narayanan

Mentors: Lana Yarosh, Estelle Smith, Haiwei Ma

Introduction

With the rise of the internet, online health communities have emerged as a means for connecting patients. While there has been much research into the effectiveness of these online health communities for patient outcomes, the effects of expressive writing on member engagement are unexplored.

However, our work in member engagement is confounded by the very nature of online health communities, as users may end use of the site through literal death. Thus, any results on member engagement need to be controlled for user death.

About CaringBridge

CaringBridge.org is an online health community created in 1997, for the purpose of documenting and sharing health journeys with others. Each "site" is a blog of a health journey, written by either the patient themselves, or a caregiver of the patient.

Sites are made up of individual journal posts, which serve as real-time updates for the health journey, and can be shared with an audience selected by the writer, in order to maintain privacy for the patient.

This research was done in collaboration with CaringBridge.org, the Center for Spiritualty and Healing, the College of Science and Engineering, and the School of Nursing at the University of Minnesota.

The CaringBridge dataset contains de-identified information from 558,210 sites and 22,333,379 users, between June 1, 2005 and June 3, 2016. All analysis and computation took place securely on the Minnesota Supercomputing Institute's high performance computing clusters.

Methods

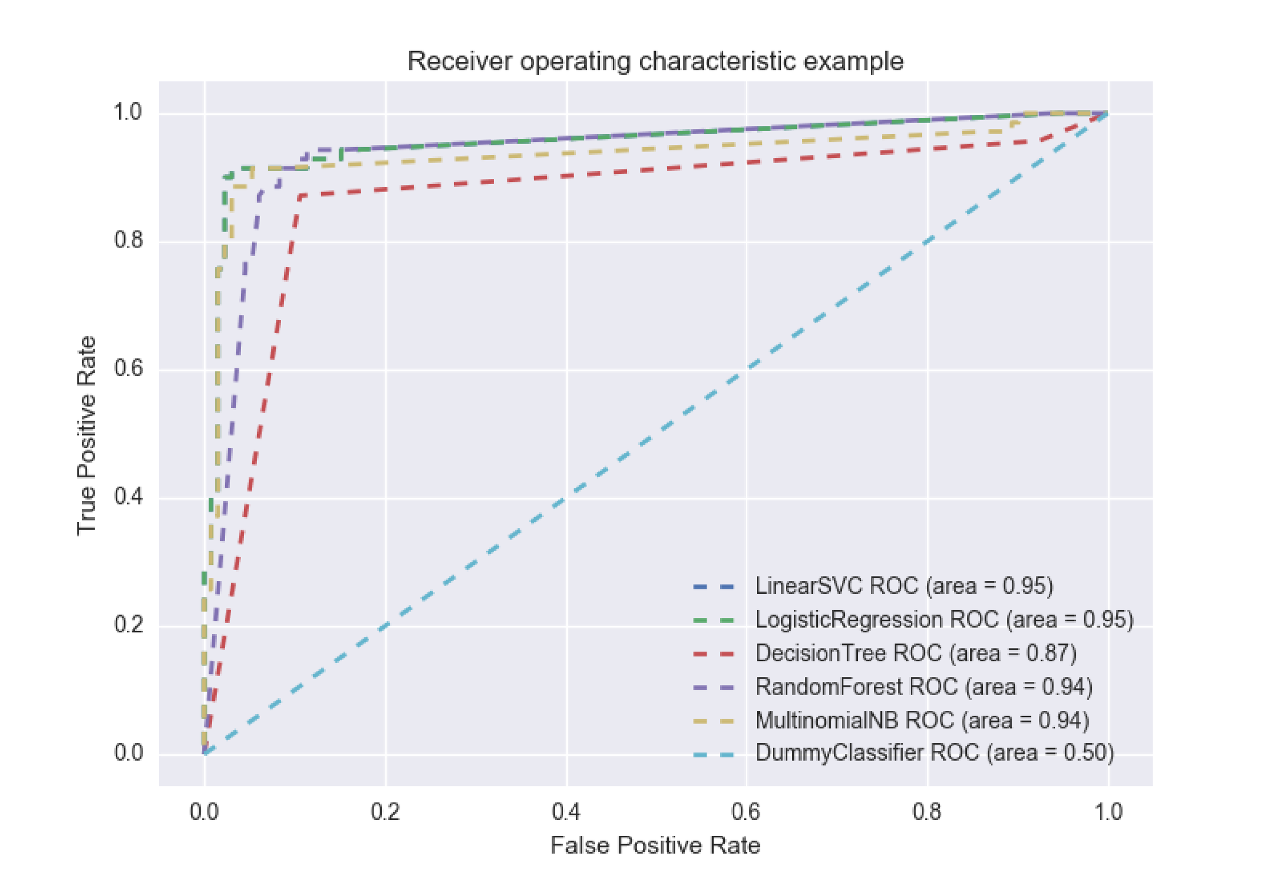
To distinguish between sites ending due to user death, and those ending for other reasons, we built a death classifier.

Our training set was composed of the last three journals from all sites self-identified as for cancer patients.

A manual set of annotations were done on a randomly selected sample of 690 sites, and each site was labeled as either deceased, or living.

For our machine learning model, we selected a list of death words, from the LIWC database, and added our own selected words and phrases to fine tune our model.

A selected sample of these words and phrases can be viewed below. Interestingly, due to the nature of the site, more obvious descriptions of death were not used by users, such as "kill*", or "death". Instead users prefer to use more metaphoric phrasings, such as "celebrate [their] life" or "with [god]", to signify the passing of a user.



Using Scikit-learn, we evaluated the performances of 5 different machine learning algorithms, on our dataset using cross-validation. For our final results, we selected Random Forest, due to the high performance, ease of training, and understandable feature importances.

LIWC Unigrams	Non-LIWC n-Grams
funeral*	passed away
death*	[memorial] service
griev*	celebrate [their] life
cemet*	in heaven
grave*	with [god]
kill*	lieu of flowers
ashes	lost [their] battle

Table 1: List of selected unigrams and n-grams from the LIWC dictionary, or selected by the research group.

Results

We found that incorporating our death classifier significantly improves results of our findings on member engagement.

Adding 128 visitors to a site increases site survival by 18.1% if we incorporate death; dropping to 11.6% without incorporating death. Likewise, adding an additional 0.85 messages/visitor drops from 9.0% to 7.6%.

Without controlling for user death, most of our correlations are underestimated, showing the importance of death in member engagement for online health communities.